Capstone Project -3

Bank Marketing Effectiveness Prediction

Done By :-

Vinit Ladse
Gaurav Bhakte
Pratiksha Kharode

Content

- Problem Statement
- Data Summary
- Exploratory Data Analysis
- Model Implementation
- Evalution Matrics
- Challenges
- Conclusion

Problem Statement

The data is related with direct marketing campaigns of a Portuguese banking institution.The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

The classification goal is to predict if the client will subscribe a term deposit (variable 'y').

<u>Data Summary</u>

The Dataset contains 17 Features with 45211 observation.

Categorical Features
- Marital - (Married , Single , Divorced)
- Job - (Management,BlueCollar,retired etc)
- Contact - (Telephone,Cellular,Unknown)
- Education - (Primary,Secondary,Tertiary)
- Month - (Jan,Feb,Mar,Apr,May etc)
- Poutcome - (Success,Failure,Other,Unknown)
- Housing - (Yes/No)
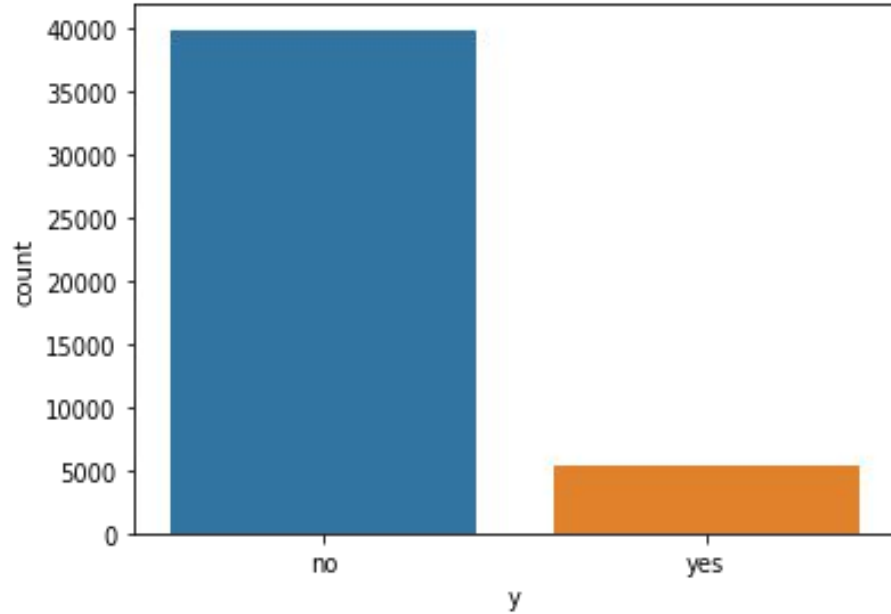- Loan - (Yes/No)
- Default - (Yes/No)

Desired target
- y - has the client subscribed a term deposit?
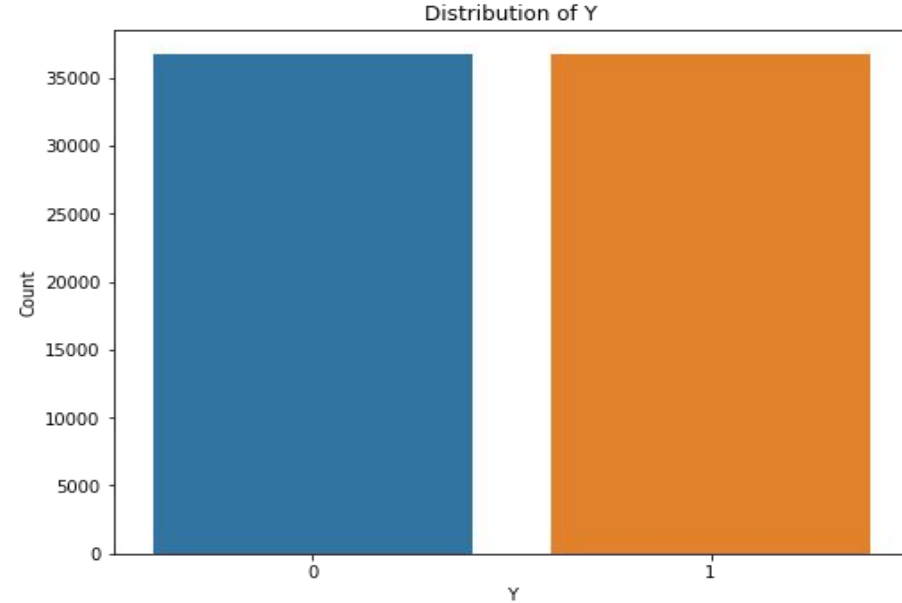 (binary: 'yes','no')

Numerical Features
- Age
- Balance
- Day
- Duration
- Campaign
- Pdays
- Previous
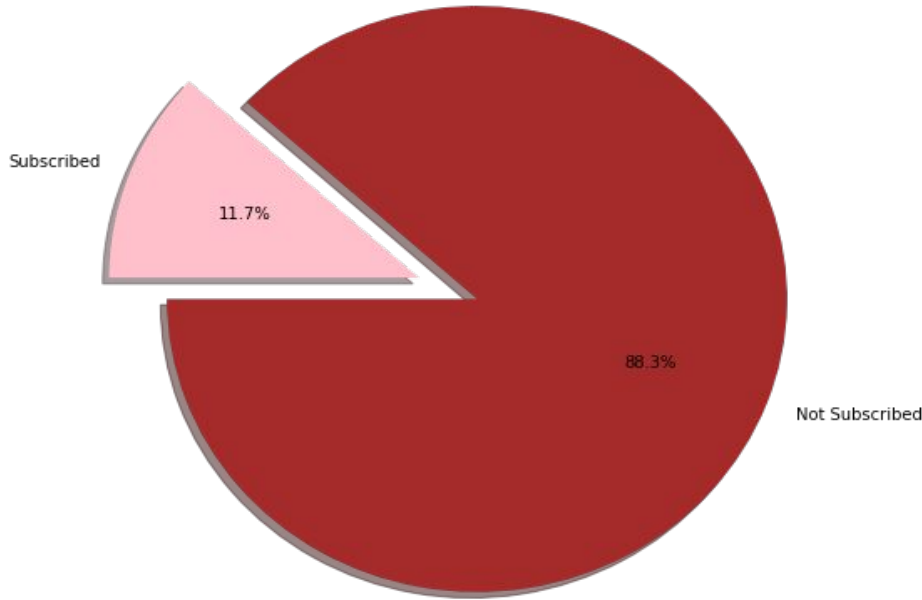
Exploratory Data Analysis (Target)
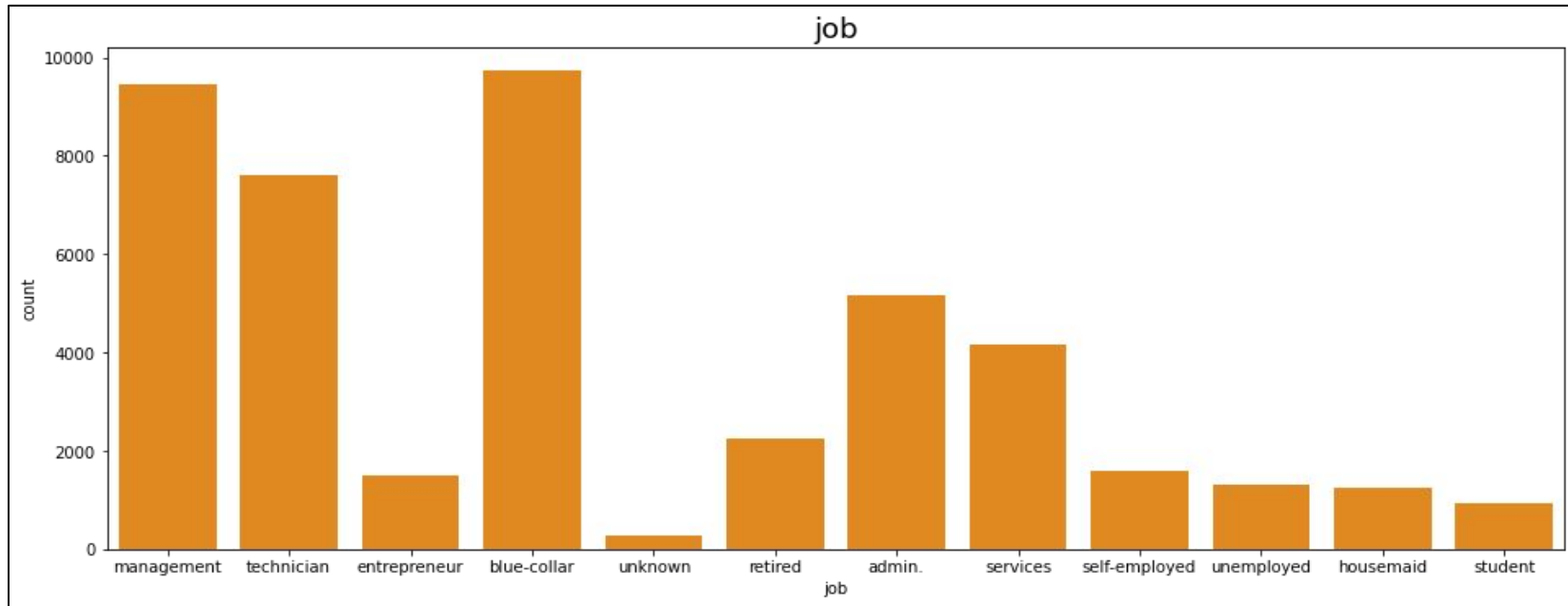
:– After

:– Before

EDA(Continued…)

How many people have subscribed the product?
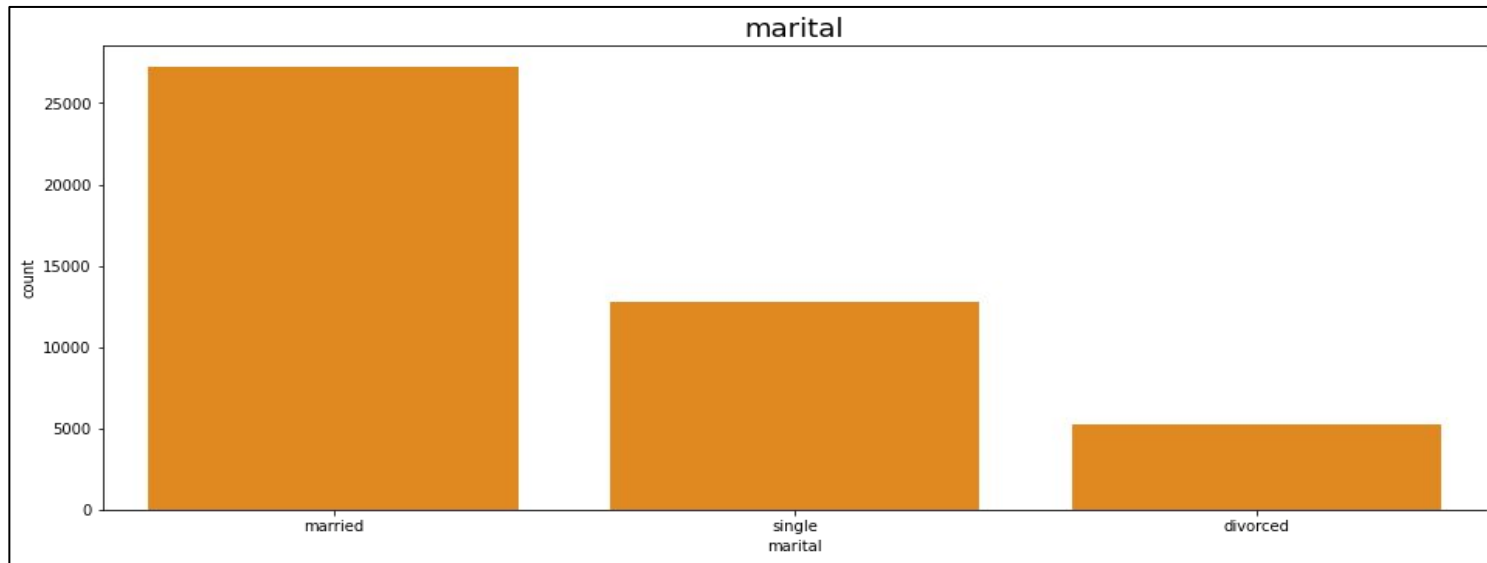
Subscribed — 11.7%

Not Subscribed — 88.3%

- From this data we can see that 88% customers did not subscribed for Term deposit

- We can say that the percentage of people subscribing to the term deposit is quite low, thus creating an imbalance in the data.

EDA(Continued…)

- Most of the customers have jobs as "management", "blue-collar" and "technician".
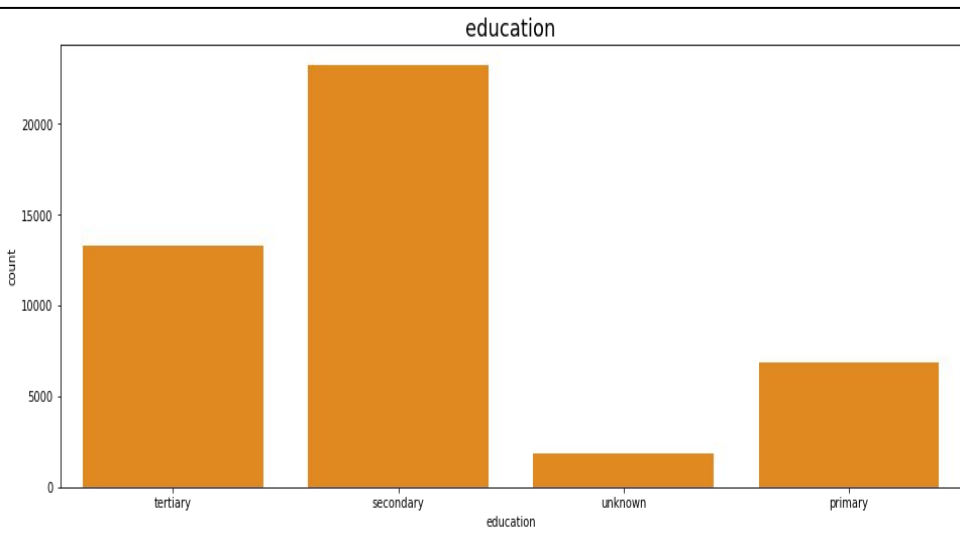- People with management jobs have subscribed more for the deposits.
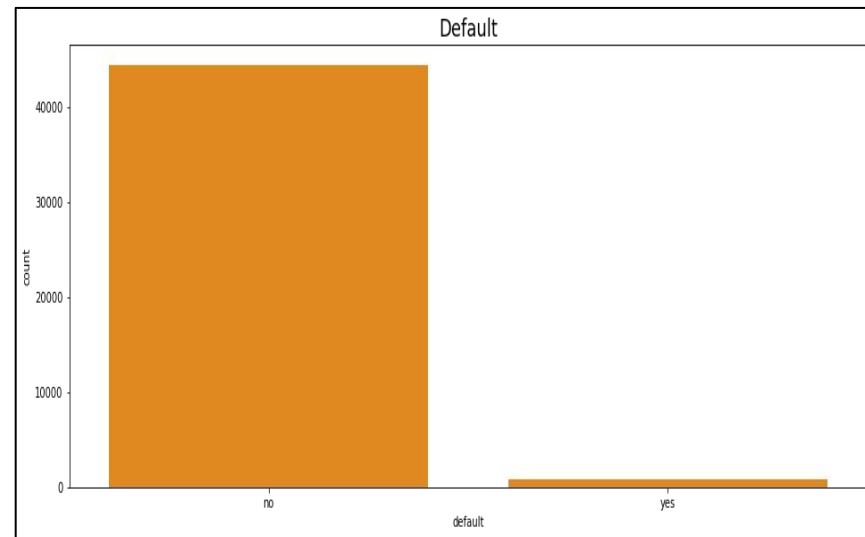
EDA(Continued…)

- Client who married are high in records.
- People who are married have subscribed for deposits more than people with any other marital status.
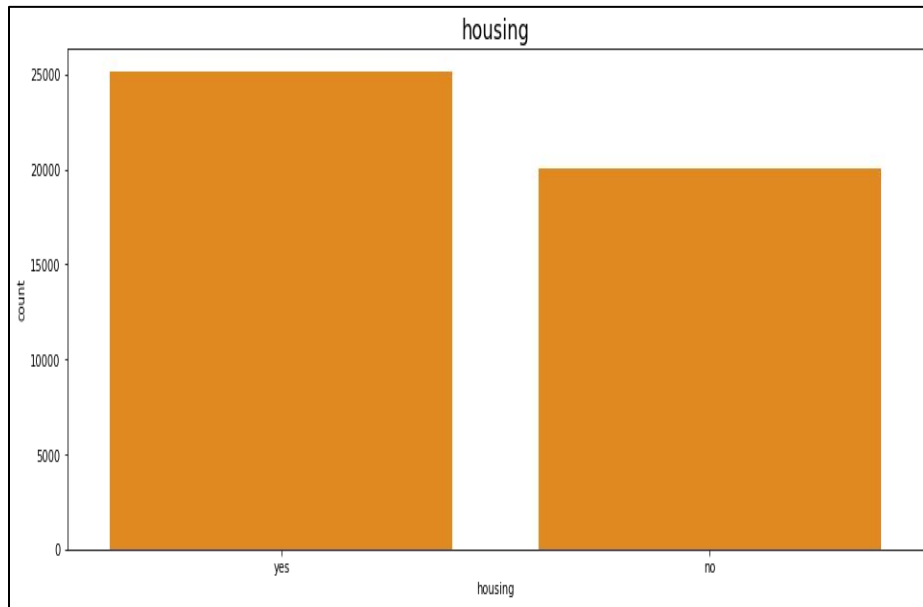
EDA(Continued…)

- Client whose education background is secondary are in high numbers.
- People with Secondary education qualification are the most who have subscribed for the deposits.
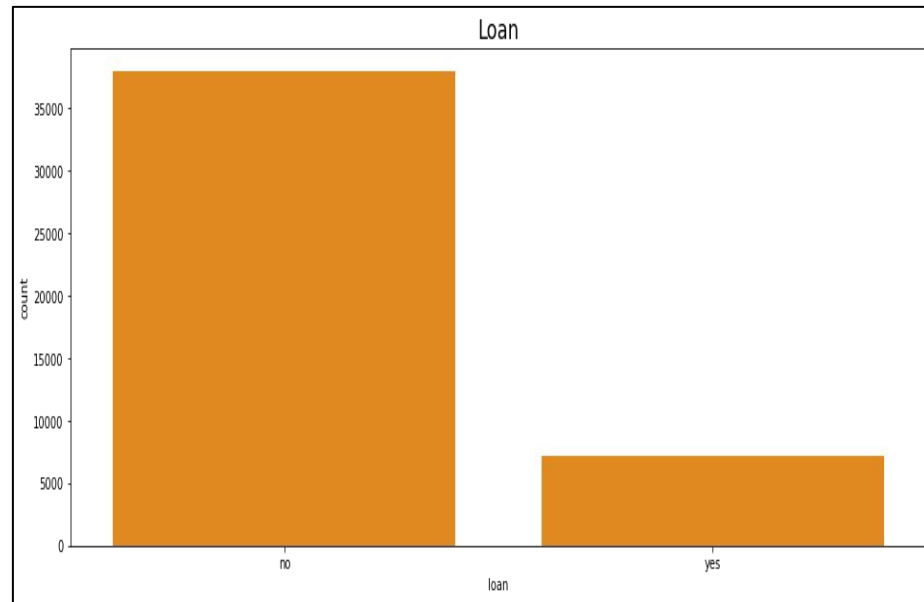
- Default feature seems to be does not play important role.
- People with default status as no are the most ones who have and have not subscribed for bank deposits.
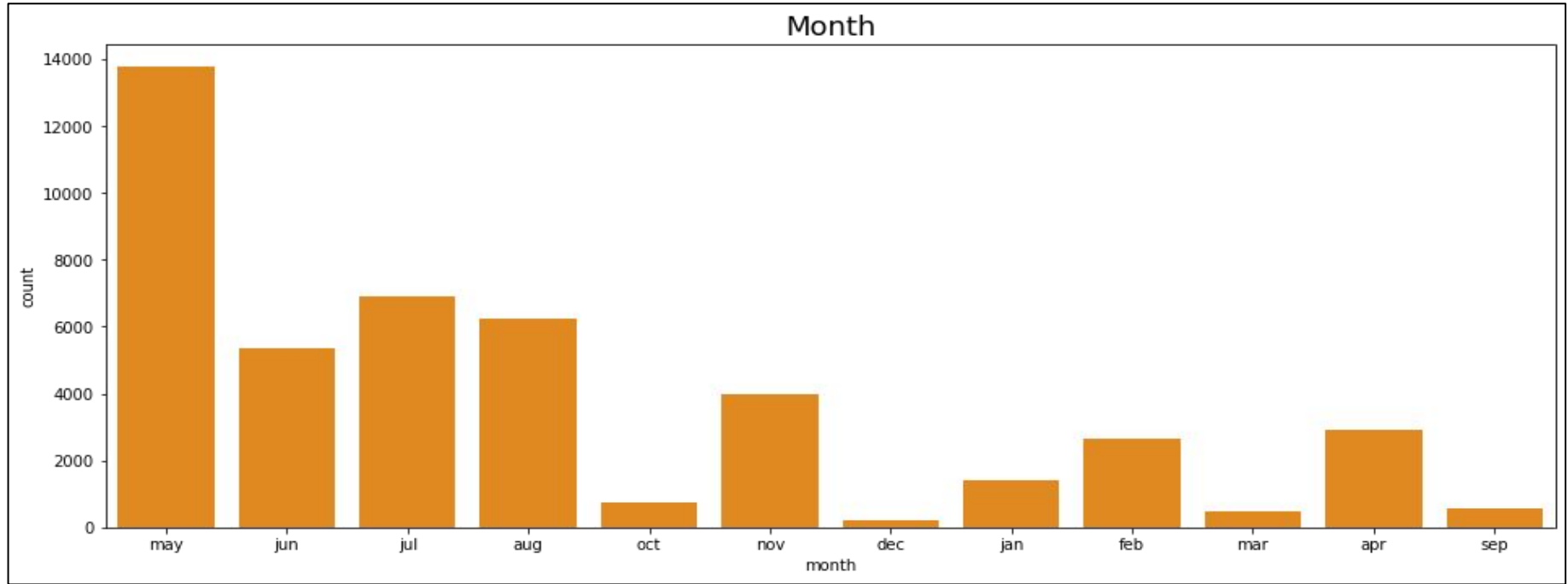
EDA(Continued…)

- People with housing loan are the most ones who have been contacted by the bank.
- People with no housing loan are the most ones who have subscribed for deposits

- People with no personal loan are the most ones who have been contacted by the bank for the deposits.
- People with no personal loan are the most ones who have not subscribed and are also the most ones who have subscribed for the deposits
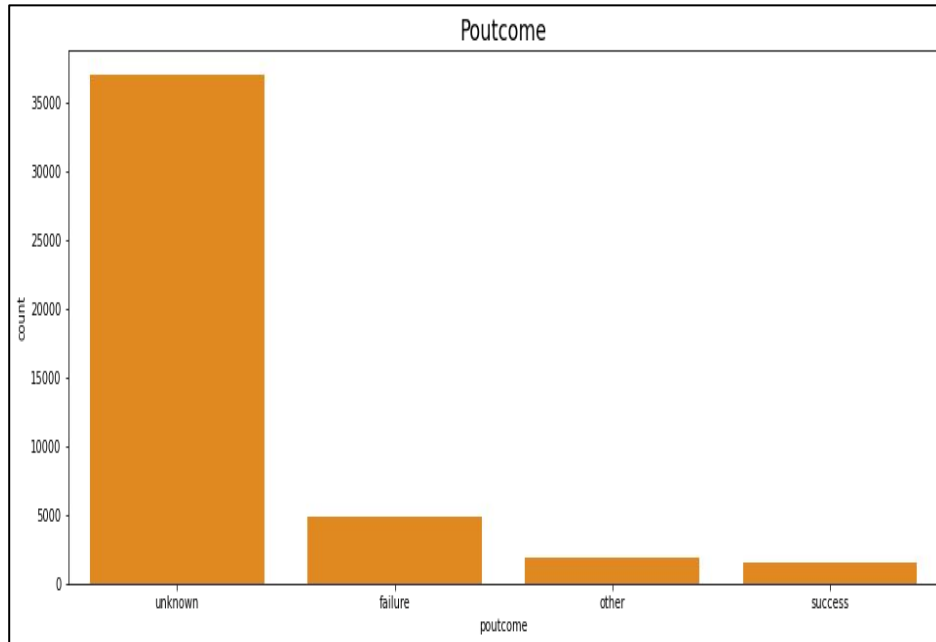
EDA(Continued…)

- Data in month of may is high and less in Dec.
- The month of the highest level of marketing activity was the month of May.
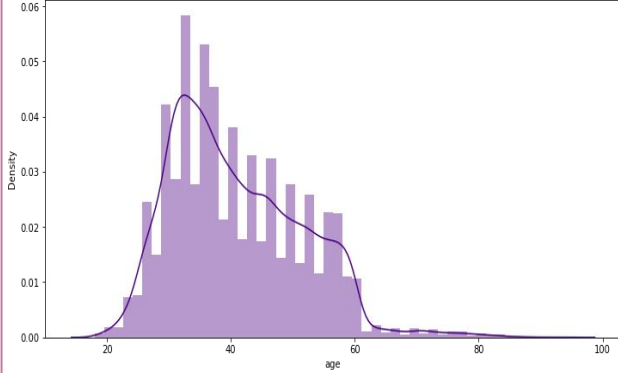
EDA(Continued…)
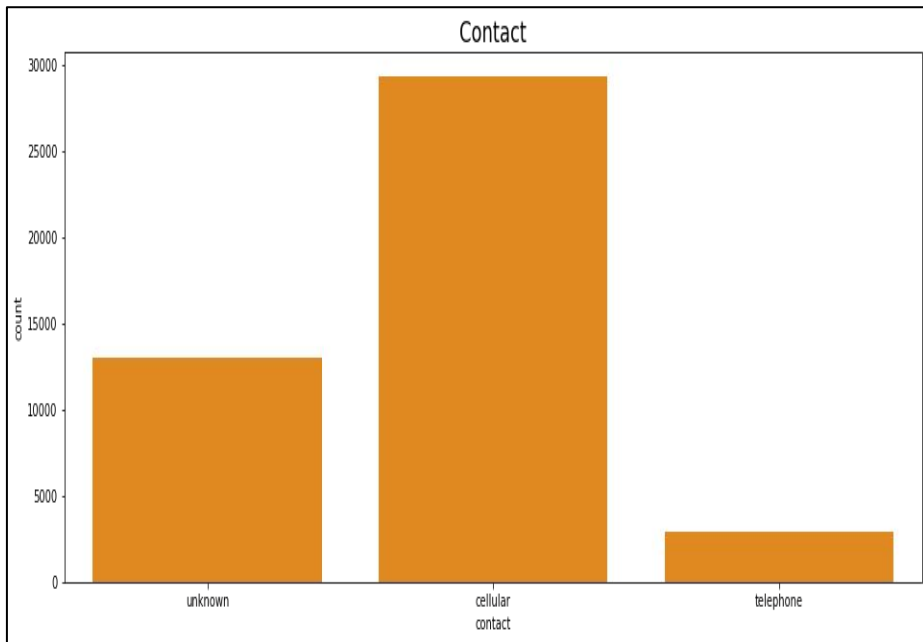
- Most people are contacted more in cellular than telephone.
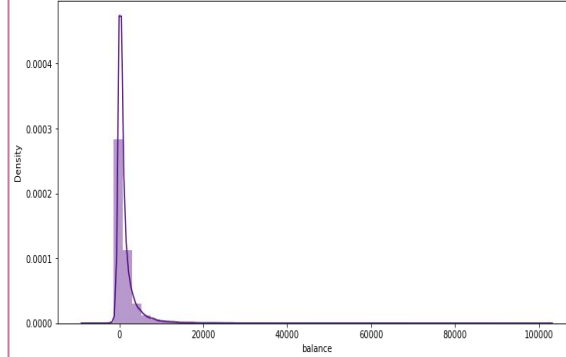- More people contacted on cellular by bank have subscribed the deposits.

- Majority of the outcome of the previous campaign is Non-Existent.
- People whose previous outcome is non-existent have actually subscribed more.
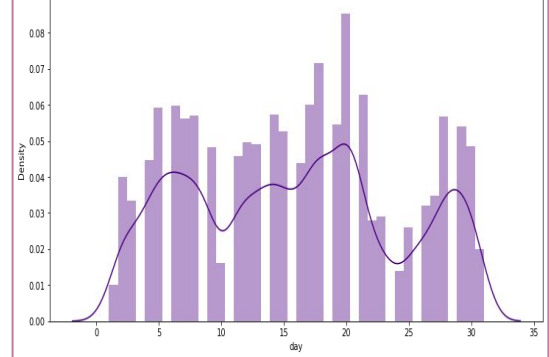
EDA(Continued..)
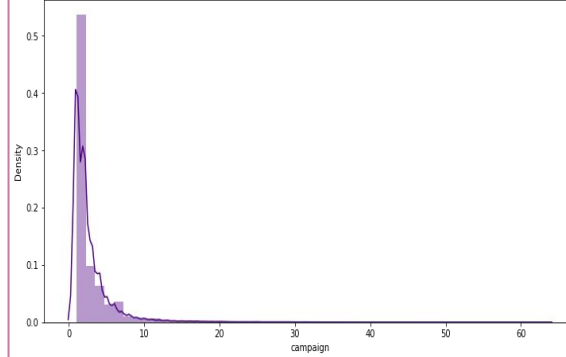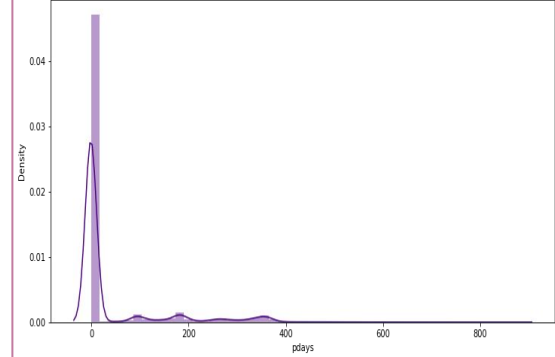
EDA(Continued…)
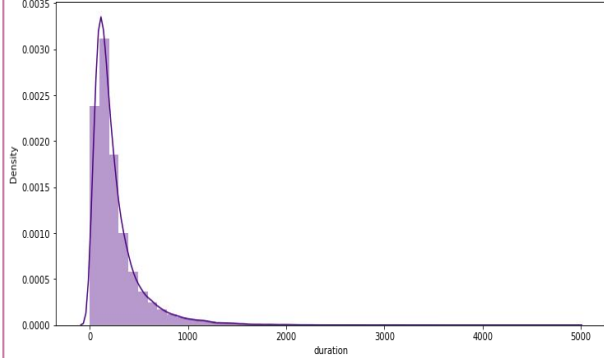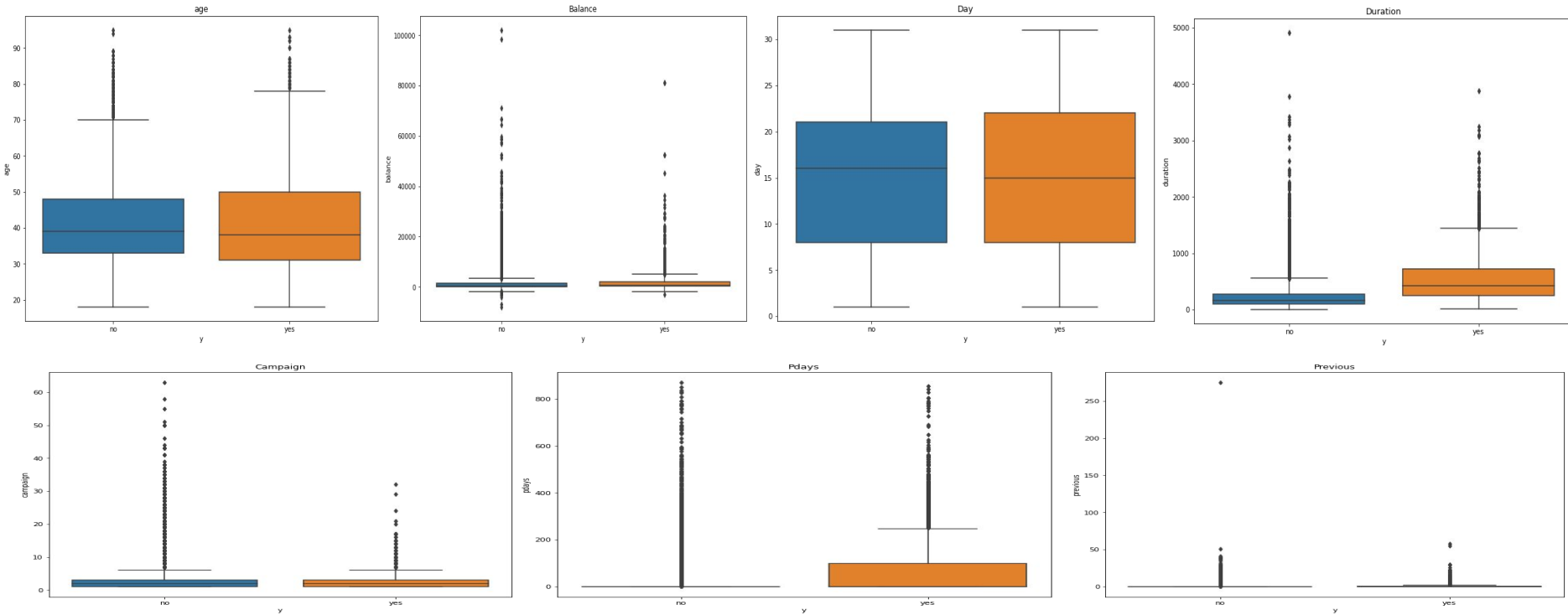

- Client shows interest on deposit who had discussion for longer duration.
- Both the customers that subscribed or didn't subscribe a term deposit, has a median age of around 38-40
- Very people has been contacted by the bank and number of days passed for previous campaign is between 0–100
- average subscription rate is below 50% if the number of contacts during the campaign exceeds 4.

Correlation
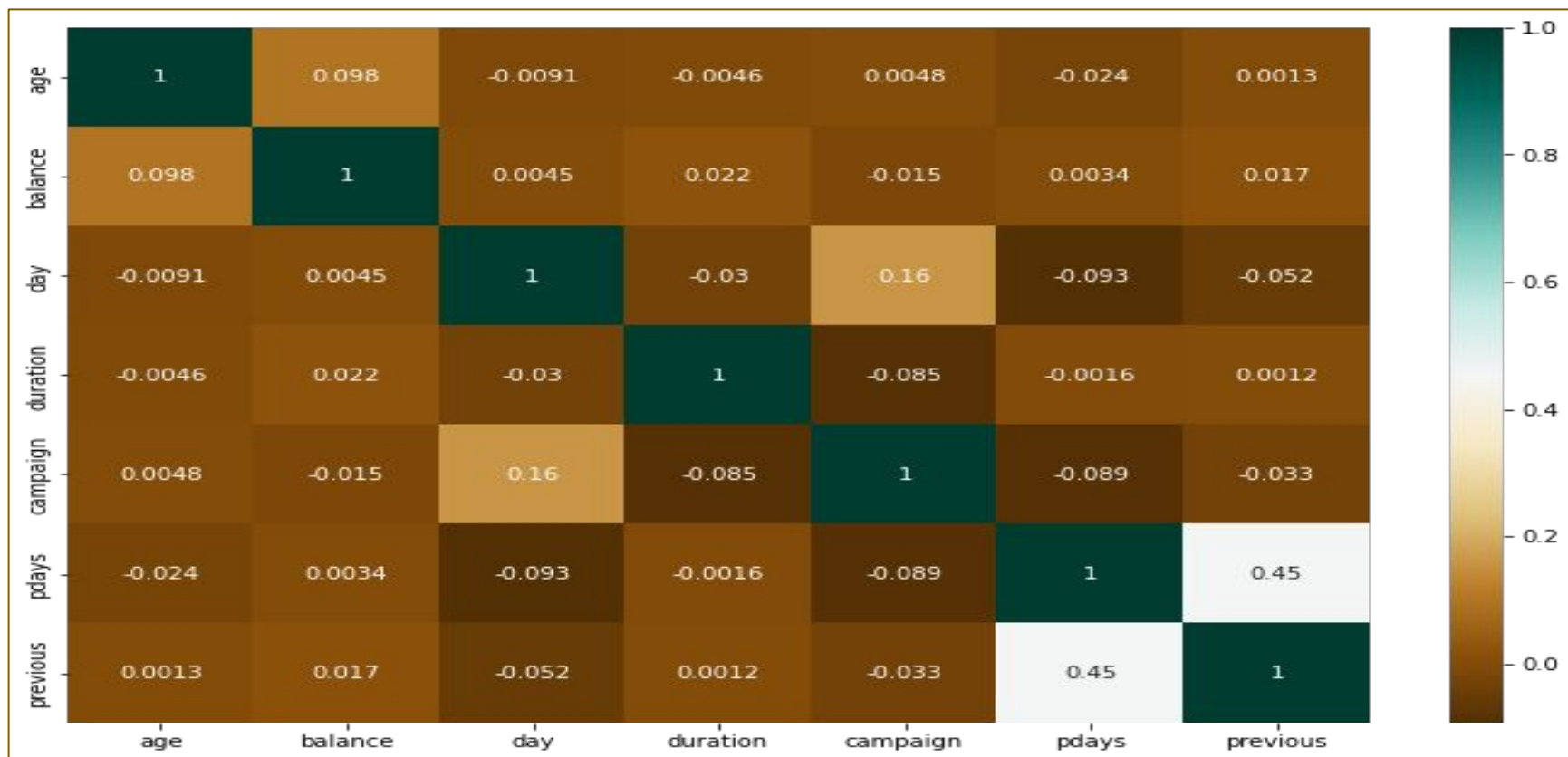
Model Implementation

Logistic Regression :-

ROC-AUC Score :-

Best Parameter :-

c : 0.1

Train Data	0.93
Test Data	0.92

Decision Tree :-

Best Parameter :-

Mean_Sample_Leaf **:** 10

Max_Depth **:** 9

Mean_Sample_split **:** 20

Train Data	0.92
Test Data	0.90

Decision Tree Features Imporatance

features importance
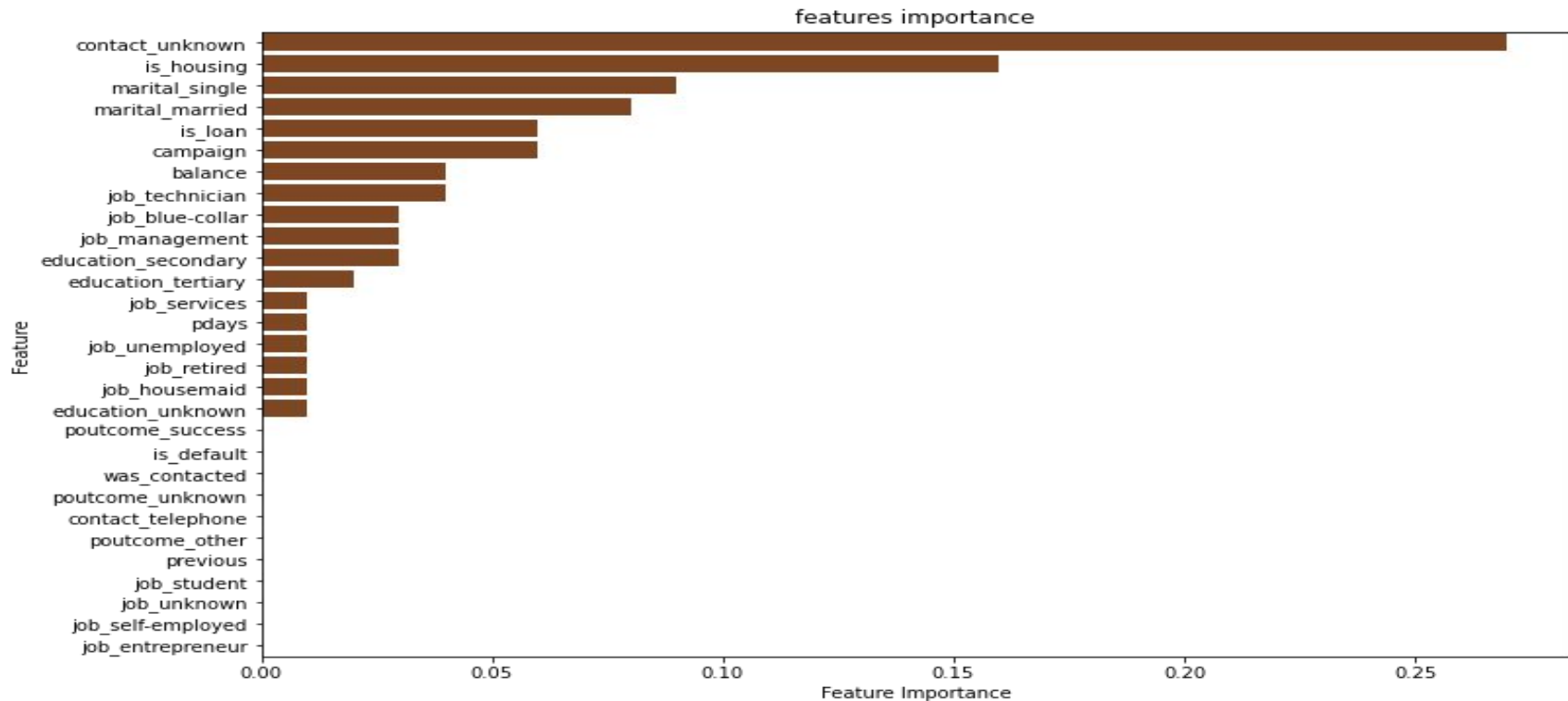
XGBoost Classifier :-

Best Parameter :-
Learning rate **:** 0.5
Max_Depth **:** 9
N_estimators **:** 125

Train Data	0.95
Test Data	0.90

XGBoost Features Imporatance

features importance
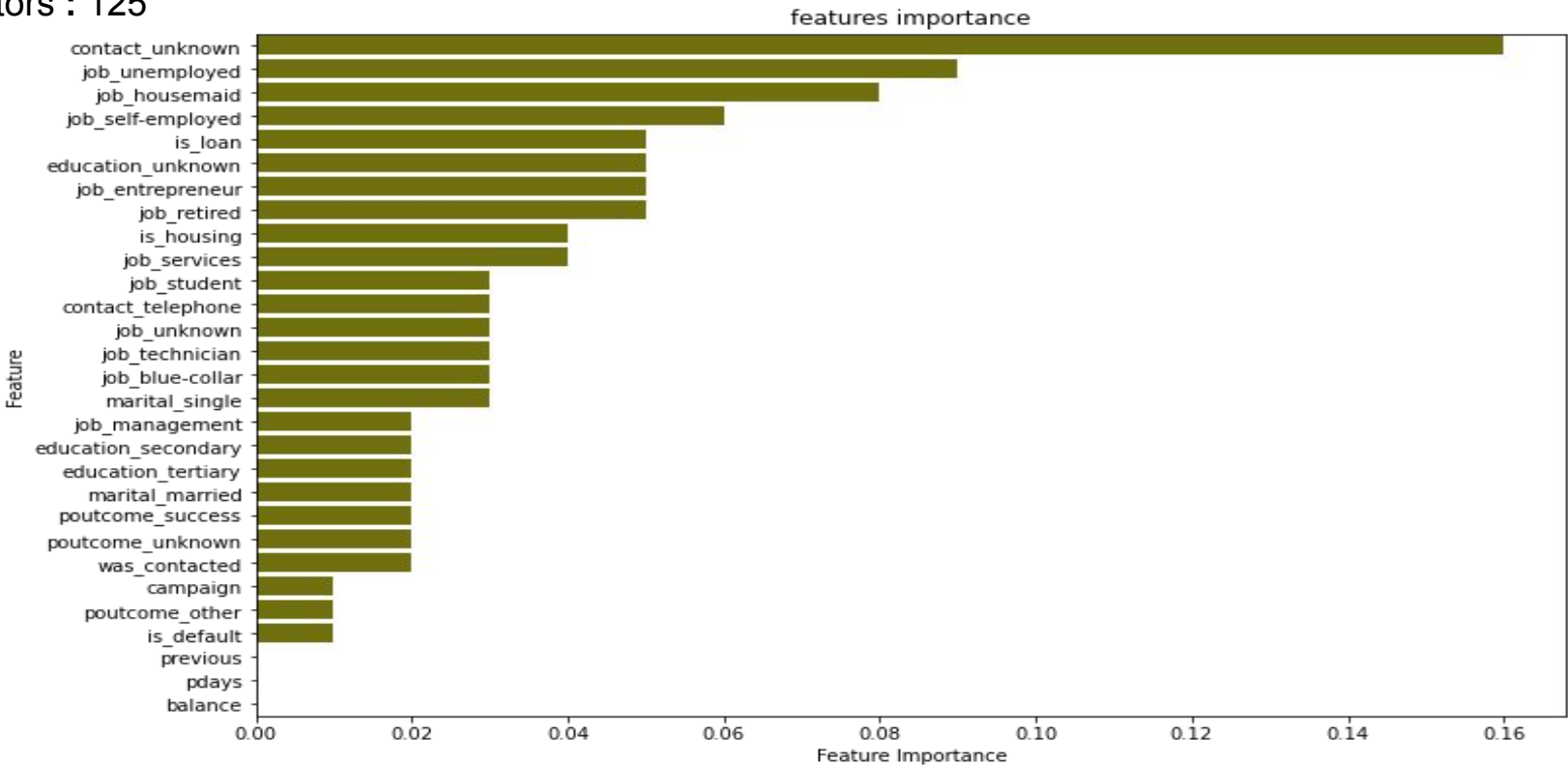
K-Nearest Neighbor:-

Best Parameter :-

n_neighbors **:** 27

ROC-AUC Score :-

Train Data	0.95
Test Data	0.93

Hyperparameter Tuning Evaluation

Model	Test AUC	Test Accuracy	F1-score	Precision
Logistic Regression	0.92	0.86	0.87	0.89
Decision Trees	0.90	0.83	0.84	0.85
XGBoost	0.93	0.90	0.91	0.92
K-NN	0.93	0.88	0.88	0.91

Challenges

- Handling Imbalanced Dataset

- Feature Engineering

- Optimising The Model

Conclusion

- For age, most of the customers are in the age range of 30-40.
- For balance, above 1000$ is like to subscribe a term deposite .
- The model can help to classify the customers on the basis on which they deposit or not
- The model helps to target the right customer rather than wasting time on wrong customer
- Comparing to all algorithms XGboost algorithm has best accuracy score and ROC-AUC score . So it is concluded as optimal model.

THANK YOU